DAHLGREN DIVISION
NAVAL SURFACE WARFARE CENTER
Dahlgren, Virginia 22448-5100

DERIVATION OF RECURSIVE UPDATE EQUATIONS FOR JOINT REPRESENTATION MIXTURE MODELS

BY GEORGE W. ROGERS AND RICHARD A. LOREY
SYSTEMS RESEARCH AND TECHNOLOGY DEPARTMENT

SEPTEMBER 1995

19960607 122

DTIC QUALITY INSPECTED 3

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE
COPY FURNISHED TO DTIC
CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO
NOT REPRODUCE LEGIBLY.

REPORT DOCUMENTATION PAGE		Form Approved OBM No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, search existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services. Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1995	3. REPORT TYPE AND DATES COVERED
4. TITLE AND SUBTITLE Derivation of Recursive Update Equations for Joint Representation Mixture Models		**5. FUNDING NUMBERS**
6. AUTHOR(s) George W. Rogers and Richard A. Lorey		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Surface Warfare Center Dahlgren Division (Code B10) 17320 Dahlgren Rd. Dahlgren, VA 22448-5100		**8. PERFORMING ORGANIZATION REPORT NUMBER** NSWCDD/TR-95/122
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	**12b. DISTRIBUTION CODE**

13. ABSTRACT (Maximum 200 words)

The joint representation mixture model is defined, likelihood functions corresponding to different levels of data categorization with respect to class are presented along with the iterative expectation-maximization (E-M) equations, and the resultant recursive E-M equations are derived.

DTIC QUALITY INSPECTED 3

14. SUBJECT TERMS Density Estimation, Mixture Models, Expectation-Maximization			**15. NUMBER OF PAGES**
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	**18. SECURITY CLASSIFICATION OF THIS PAGE** UNCLASSIFIED	**19. SECURITY CLASSIFICATION OF ABSTRACT** UNCLASSIFIED	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500　DTIC QUALITY INSPECTED 3　i/ii

Standard Form 298 (Rev 2-89)
Prescribed by ANSI std. Z39-18
298-102

FOREWORD

Density estimation plays a central role in probabilistic pattern recognition and signal processing. As data sets get larger, the cost of identifying a definitive class with each observation can become prohibitive. Instead, it becomes important to develop ways to process the data in ways that make use of all available information.

This work was done under the joint support of the NSWCDD Independent Research program and the Office of Naval Research (R&T #4424314).

This report was reviewed by Dr. Richard Lorey, Head, Advanced Computation Technology Group, and Dr. Glenn Moore, Head, Naval Surface Warfare Center Independent Research Program.

Approved by:

Mary C. Lacey

MARY E. LACEY, Head
Systems Research and Technology Department

CONTENTS

INTRODUCTION

Finite mixture models have proven to be quite flexible as parametric probability density function estimators.[1,2] Recently an adaptive mixture model was presented whose complexity or number of terms is determined in a data driven manner.[3] This approach has made possible the use of mixture models within a semiparametric setting, and thus of much more general applicability/ utility than was possible under rigid parametric assumptions.

This semiparametric use of mixture models has resulted in efforts to develop alternative adaptive mixture model algorithms.[4,5] Recent applications of semiparametric mixture model density estimation can be found in References 6 through 11. Thus, in addition to the traditional parametric uses of mixture models, the semiparametric application of mixture models is now well established.

One of the problems that arises in many applications of mixture models to density estimation of large scale data sets is that, as the size of the data set increases, the class labeled data becomes a (small) subset of the total data set; that is, while many small data sets may have all the observations labeled as to class membership, large data sets often consist of labeled subsets plus a potentially large unlabeled subset.

The reason for this can be illustrated with an image processing example. Suppose that features are to be computed for each pixel for a number of images and that densities are to be computed for each class. Depending on the problem, the classes may correspond to vehicles, buildings, woods, and open terrain, or to tumorous and nontumorous tissue. If all the available data is to be used, the work in allocating each original pixel to one of the classes can easily become prohibitive. The more usual case is that only a representative subset of training data are class labeled with the balance either uncategorized or partially categorized. An example of the latter case is that it may be easy to say that there are no vehicles in this image, no buildings in that one, and so on but very difficult or time consuming to identify each pixel corresponding to each class in each image. It is often the case in medical imagery that ground truth cannot be established definitively without a biopsy, again leading to less than full categorization of the observations.

Thus it is desirable to have a unified framework for handling this combined supervised (class labeled data)/unsupervised (unlabeled data) problem. This was the motivation behind the development of joint representation mixture models.[12] When dealing with large data sets, by which we mean 100,000 observations or more, the iterative expectation-maximization (E-M) equations often become impractical. One method of dealing with this much data is to go to recursive formulations of the E-M equations. This approach also makes possible the implementation of the adaptive mixure model approach.[3] The derivation of the recursive E-M equations for joint representation mixture models is the focus of this report.

JOINT REPRESENTATION MIXTURE MODELS

FINITE MIXTURE MODELS

Given a probability density function that can be represented as a finite (g term) mixture model

$$p(x|\psi) = \sum_{i=1}^{g} \pi_i f(x;\theta_i),$$ (1)

where $f(\bullet|\theta)$ denotes a generic member of the chosen parametric family, the likelihood function for n observations is given by

$$L(\psi) = \prod_{j=1}^{n} \sum_{i=1}^{g} \pi_i f(x_j;\theta_i).$$ (2)

The vector θ_i represents the parameter set for the ith mixture component while ψ represents the combined total parameter set including the mixing coefficients π_i. The log-likelihood function is

$$lnL(\psi) = \sum_{j=1}^{n} ln\left[\sum_{i=1}^{g} \pi_i f(x_j;\theta_i)\right].$$ (3)

The maximum likelihood update equations can be obtained by taking derivatives of the log-likelihood function with respect to the mixture model parameters, setting the resulting expressions equal to zero, and solving for the parameters.

JOINT REPRESENTATION MIXTURE MODELS

Consider the Joint Representation Mixture Model defined by

$$p(x|\psi) = \sum_{i=1}^{g} p(\text{term } i) p(x|\text{term } i) \sum_{m=1}^{M} p(\text{class } m|\text{term } i) = \sum_{i=1}^{g} \pi_i f(x;\theta_i) \sum_{m=1}^{M} \zeta_{im}$$ (4)

where $\zeta_{im} = p(\text{class } m|\text{term } i)$ is an intra-term class mixing coefficient that gives the relative proportion of the ith term associated with the mth class with the constraint

$$\sum_{m=1}^{M} \zeta_{im} = \sum_{m=1}^{M} p(\text{class } m|\text{term } i) = 1.$$ (5)

This constraint merely says that for each term independently, the class mixing coefficients must sum to one, or equivalently, that an observation from term i must belong to one of the M classes with probability one.

The mixture model defined in Equations (4) and (5) represents a significant departure from traditional mixture model usage. Historically a single mixture model has been used for either performing unsupervised clustering or to generate a probability density function for observations from a single class. If observations from multiple classes are to be dealt with, then a separate mixture model is developed for data from each class. This latter approach leaves open the question of how to incorporate partially (class) categorized or uncategorized observations when there are separate mixture models for each class. As will be seen, the joint representation formulation leads to a unified treatment of these cases.

JOINT REPRESENTATION MIXTURE MODEL LIKELIHOOD FUNCTIONS

The likelihood function for class categorized data is

$$
\begin{aligned}
L_c(\psi) &= \prod_{j=1}^{n_c} \prod_{m=1}^{M} \left\{ \left[p\left(x_j \middle| (\psi \cap \text{Class } m)\right) \right]^{z_{jm}} \right\} \\
&= \prod_{j=1}^{n_c} \prod_{m=1}^{M} \left\{ \left[\sum_{i=1}^{g} \zeta_{im} \pi_i f(x_j;\theta_i) \right]^{z_{jm}} \right\}.
\end{aligned}
\tag{6}
$$

Here, z_{jm} is a binary valued class indicator function. For observation j from class h, $z_{jh} = 1$ and $z_{jm} = 0$ for $m \neq h$. It thus can be considered as a picking function. It is used to *pick out* the desired contribution to the likelihood function. For each term in the product where it has the value zero, the contribution to the product is one so that the likelihood is unaffected. The log-likelihood function for class categorized data can be written

$$
\begin{aligned}
lnL_c(\psi) &= \sum_{j=1}^{n_c} \sum_{m=1}^{M} ln \left\{ \left[\sum_{i=1}^{g} \zeta_{im} \pi_i f(x_j;\theta_i) \right]^{z_{jm}} \right\} \\
&= \sum_{j=1}^{n_c} \sum_{m=1}^{M} z_{jm} ln \left\{ \sum_{i=1}^{g} \zeta_{im} \pi_i f(x_j;\theta_i) \right\} \\
&= \sum_{j=1}^{n_c} ln \sum_{i=1}^{g} \sum_{m=1}^{M} z_{jm} \zeta_{im} \left[\pi_i f(x_j;\theta_i) \right].
\end{aligned}
\tag{7}
$$

To write down the likelihood for partially (class) categorized data, first consider the likelihood appropriate for the case where the data is both class and term categorized. In this case, the

likelihood is

$$L_{ct}(\psi) = \prod_{j=1}^{n_c} \prod_{m=1}^{M} \prod_{i=1}^{g} \{ [\zeta_{im}\pi_i f(x_j;\theta_i)]^{z_{jm}z_{ji}} \} .$$

(8)

Where as before, z_{jm} (z_{ji}) is a binary valued class (term) indicator function. For observation j from class h, $z_{jh} = 1$ and $z_{jm} = 0$ for $m \neq h$, while for observation j from term k, $z_{jk} = 1$ and $z_{ji} = 0$ for $i \neq k$.

In the absence of complete knowledge of z_{jm} and/or z_{ji} the usual procedure is to use expected values for either/both as

$$E(z_{jm}) = \xi_{jm}$$

(9)

and

$$E(z_{ji}) = \tau_{ij}^p .$$

(10)

Then for partially (class) categorized data, the likelihood is

$$L_p(\psi) = \prod_{j=1}^{n_p} \prod_{i=1}^{g} \prod_{m=1}^{M} \{ [\zeta_{im}\pi_i f(x_j;\theta_i)]^{\xi_{jm}\tau_{ij}^p} \}$$

(11)

whence

$$lnL_p(\psi) = \sum_{j=1}^{n_p} \sum_{i=1}^{g} \sum_{m=1}^{M} \{ \xi_{jm}\tau_{ij}^p ln[\zeta_{im}\pi_i f(x_j;\theta_i)] \}$$

$$= \sum_{j=1}^{n_p} \sum_{i=1}^{g} \sum_{m=1}^{M} ln \{ [\zeta_{im}\pi_i f(x_j;\theta_i)]^{\xi_{jm}\tau_{ij}^p} \}$$

(12)

where ξ_{jm} is a prior or expected probability of class membership with

$$\sum_{m=1}^{M} [\xi_{jm}] = 1 ,$$

(13)

and

$$\tau_{ij}^p = \frac{\pi_i f(x_j;\theta_i)}{\sum_{i=1}^{g} \pi_i f(x_j;\theta_i)}$$

(14)

is the expectation (posterior probability) that the jth observation came from the ith mixture term.

Since this is an expectation, it will be held fixed while taking derivatives. Two common methods for specifying partial categorization are (1) to give prior probabilities for each class for observation j, or (2) to specify that the priors are zero for some subset of the classes and to use posterior probabilities across the remaining possible classes for the *unknown* ξ_{jm}.

For uncategorized data,

$$L_u(\psi) = \prod_{j=1}^{n_u} \left[\sum_{i=1}^{g} \sum_{m=1}^{M} \zeta_{im} \pi_i f(x_j; \theta_i) \right] = \prod_{j=1}^{n_u} \sum_{i=1}^{g} [\pi_i f(x_j; \theta_i)] , \qquad (15)$$

$$lnL_u(\psi) = \sum_{j=1}^{n_u} ln \sum_{i=1}^{g} \sum_{m=1}^{M} \zeta_{im} \pi_i f(x_j; \theta_i) = \sum_{j=1}^{n_u} ln \sum_{i=1}^{g} [\pi_i f(x_j; \theta_i)] . \qquad (16)$$

For combined categorized/partially categorized/uncategorized data,

$$lnL(\psi) = lnL_c(\psi) + lnL_p(\psi) + lnL_u(\psi) \qquad (17)$$

or

$$lnL(\psi) = \sum_{j=1}^{n_c} ln \sum_{i=1}^{g} \sum_{m=1}^{M} z_{jm} \zeta_{im} [\pi_i f(x_j; \theta_i)]$$
$$+ \sum_{j=1}^{n_p} \sum_{m=1}^{M} \sum_{i=1}^{g} \xi_{jm} \tau_{ij} ln [\zeta_{im} \pi_i f(x_j; \theta_i)] + \sum_{j=1}^{n_u} ln \sum_{i=1}^{g} \pi_i f(x_j; \theta_i) . \qquad (18)$$

Historically with mixture models, reference to categorization of data has been with respect to which term of the mixture model the observation is associated. While this is logical when each term is ascribed a *class* status as in clustering, in this work, a completely different definition of categorized data is being used. In this case, the concern is that of categorizing data only with respect to class membership rather than with respect to individual mixture model terms.

To derive the maximum likelihood update equations, the parameter values that give a maximum of the log-likelihood function must be found. This can be accomplished by taking the derivative with respect to each parameter, setting it equal to zero, and solving the resultant system of equations for the parameters. This has been done previously[12] so that those results will be taken as the starting point in developing the recursive versions here.

MAXIMUM LIKELIHOOD E-M EQUATIONS

The joint representation E-M equations are[12]

$$\zeta_{im} = \left[\frac{1}{\sum\limits_{j=1}^{n_c} \sum\limits_{m=1}^{M} \tau_{ijm} + \sum\limits_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\sum\limits_{j=1}^{n_c} \tau_{ijm} + \sum\limits_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p \right], \tag{19}$$

$$\pi_i = \frac{1}{(n_c + n_p + n_u)} \left[\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u \right], \tag{20}$$

$$\mu_i = \frac{\sum\limits_{j=1}^{n_c} \left[\tau_{ij}^c x_j \right] + \sum\limits_{j=1}^{n_p} \left[\tau_{ij}^p x_j \right] + \sum\limits_{j=1}^{n_u} \left[\tau_{ij}^u x_j \right]}{\left[\sum\limits_{j=1}^{n_c} \left[\tau_{ij}^c \right] + \sum\limits_{j=1}^{n_p} \left[\tau_{ij}^p \right] + \sum\limits_{j=1}^{n_u} \left[\tau_{ij}^u \right] \right]}, \tag{21}$$

$$\Sigma_{ii} = \frac{\sum\limits_{j=1}^{n_c} \tau_{ij}^c (x_j - \mu_i)^2 + \sum\limits_{j=1}^{n_p} \tau_{ij}^p (x_j - \mu_i)^2 + \sum\limits_{j=1}^{n_u} \tau_{ij}^u (x_j - \mu_i)^2}{\left[\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u \right]}, \tag{22}$$

where

$$\tau_{ijm} = \frac{z_{jm} \zeta_{im} \pi_i f(x_j; \theta_i)}{\sum\limits_{i=1}^{g} \sum\limits_{m=1}^{M} z_{jm} \zeta_{im} \pi_i f(x_j; \theta_i)}, \tag{23}$$

and

$$\tau_{ij}^c = \sum\limits_{m=1}^{M} \tau_{ijm} = \frac{\zeta_{im_j} \pi_i f(x_j; \theta_i)}{\sum\limits_{i=1}^{g} \zeta_{im_j} \pi_i f(x_j; \theta_i)} = \frac{\sum\limits_{m=1}^{M} z_{jm} \zeta_{im} \pi_i f(x_j; \theta_i)}{\sum\limits_{i=1}^{g} \sum\limits_{m=1}^{M} z_{jm} \zeta_{im} \pi_i f(x_j; \theta_i)}. \tag{24}$$

Similarly

$$\tau_{ij}^p = \tau_{ij}^u = \frac{\pi_i f(x_j;\theta_i)}{\displaystyle\sum_{i=1}^{g} \pi_i f(x_j;\theta_i)}, \qquad (25)$$

where it is to be remembered that τ_{ij}^p is only computed when x_j is a partially categorized observation and similarly for τ_{ij}^u.

The E-M algorithm then consists of iterating the expectation step consisting of evaluating Equations (23), (24), and (25) for the appropriate observations and the maximization step, which consists of evaluating new parameter values using Equations (19) through (22).

The multivariate versions can be obtained by making x_j and μ_i vector quantities and Σ_i a matrix. The equations for μ_i and Σ_i become

$$\mu_i^k = \frac{\displaystyle\sum_{j=1}^{n_c} \left[\tau_{ij}^c x_j^k\right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^p x_j^k\right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^u x_j^k\right]}{\left[\displaystyle\sum_{j=1}^{n_c} \left[\tau_{ij}^c\right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^p\right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^u\right]\right]} \qquad (26)$$

and

$$\Sigma_i^{kl} = \frac{\displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c \left(x_j^k - \mu_i^k\right)\left(x_j^l - \mu_i^l\right) + \sum_{j=1}^{n_p} \tau_{ij}^p \left(x_j^k - \mu_i^k\right)\left(x_j^l - \mu_i^l\right) + \sum_{j=1}^{n_u} \tau_{ij}^u \left(x_j^k - \mu_i^k\right)\left(x_j^l - \mu_i^l\right)}{\left[\displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u\right]}, \qquad (27)$$

where the component indices are denoted by superscripts.

Finally, once the joint representation mixture model has been obtained based on any combination of class categorized, partial class categorized, and uncategorized data, if desired the probability density function for an individual class can be obtained through

$$\hat{p}\,(x|\text{Class y},\psi)\;=\;\frac{\displaystyle\sum_{i=1}^{g}\zeta_{iy}\pi_i f(x;\theta_i)}{\displaystyle\sum_{i=1}^{g}\zeta_{iy}\pi_i}\,.\tag{28}$$

This gives a properly normalized mixture model density estimate for an individual class.

These results serve as the starting point for the derivation of the recursive update equations.

DERIVATION OF RECURSIVE UPDATE EQUATIONS FOR JOINT REPRESENTATION MODEL

Recursive Update Equation for ζ

Consider first $\zeta_{im}^{(n_c+1,\,n_p)}$, which implies that the latest observation is class categorized. In terms of the E-M expression,

$$\zeta_{im}^{(n_c+1,\,n_p)}\;=\;\left[\frac{\displaystyle\sum_{j=1}^{n_c+1}\tau_{ijm}+\sum_{j=1}^{n_p}\xi_{jm}\tau_{ij}^{p}}{\displaystyle\sum_{j=1}^{n_c+1}\tau_{ij}^{c}+\sum_{j=1}^{n_p}\tau_{ij}^{p}}\right]\tag{29}$$

or, equivalently

$$\zeta_{im}^{(n_c+1,\,n_p)}\;=\;\left[\frac{\{\tau_{i(n_c+1)m}\}+\displaystyle\sum_{j=1}^{n_c}\tau_{ijm}+\sum_{j=1}^{n_p}\xi_{jm}\tau_{ij}^{p}}{\tau_{i(n_c+1)}^{c}+\displaystyle\sum_{j=1}^{n_c}\tau_{ij}^{c}+\sum_{j=1}^{n_p}\tau_{ij}^{p}}\right]\,.\tag{30}$$

The right-hand side of this equation can be broken into two terms, corresponding to the last observation and all previous observations, respectively.

$$\zeta_{im}^{(n_c+1, n_p)} = \left[\frac{\tau_{i(n_c+1)m}}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] + T_2 \qquad (31)$$

where the second term is given by

$$T_2 = \left[\frac{\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right]. \qquad (32)$$

This can be rewritten as

$$T_2 = \left[\frac{\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\frac{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right], \qquad (33)$$

which, with a minor rearrangement of terms, becomes

$$T_2 = \left[\frac{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\frac{\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right]. \qquad (34)$$

The last term in brackets on the right-hand side is just $\zeta_{im}^{(n_c, n_p)}$, so that

$$T_2 = \left[1 - \frac{\tau_{i(n_c+1)}^c}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\zeta_{im}^{(n_c, n_p)} \right]. \qquad (35)$$

Then

$$\zeta_{im}^{(n_c+1,\, n_p)} = \zeta_{im}^{(n_c,\, n_p)} + \frac{\tau_{i(n_c+1)m} - \tau_{i(n_c+1)}^c \cdot \zeta_{im}^{(n_c,\, n_p)}}{\tau_{i(n_c+1)}^c + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}. \tag{36}$$

Using the identity

$$\tau_{i(n_c+1)m} = z_{(n_c+1)m} \cdot \sum_{m=1}^{M} \left\{ \tau_{i(n_c+1)m} \right\} = z_{(n_c+1)m} \cdot \tau_{i(n_c+1)}^c$$

the final expression can be written as

$$\zeta_{im}^{(n_c+1,\, n_p)} = \zeta_{im}^{(n_c,\, n_p)} + \frac{\tau_{i(n_c+1)}^c \cdot \left[z_{(n_c+1)m} - \zeta_{im}^{(n_c,\, n_p)} \right]}{\tau_{i(n_c+1)}^c + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}. \tag{37}$$

Similarly, for a partially categorized observation,

$$\zeta_{im}^{(n_c,\, n_p+1)} = \left[\frac{\displaystyle\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p+1} \xi_{jm} \tau_{ij}^p}{\displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p+1} \tau_{ij}^p} \right] = \left[\frac{\xi_{n_p+1,\, m} \tau_{i,\, n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p}{\tau_{i,\, n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \tag{38}$$

or, as before

$$\zeta_{im}^{(n_c,\, n_p+1)} = \left[\frac{\xi_{n_p+1,\, m} \tau_{i,\, n_p+1}}{\tau_{i,\, n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] + T_2 \tag{39}$$

with

$$T_2 = \left[\frac{\displaystyle\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm}\tau_{ij}^p}{\tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right]. \tag{40}$$

Proceeding as before in the completely categorized case,

$$T_2 = \left[\frac{\displaystyle\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm}\tau_{ij}^p}{\tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\frac{\tau_{i,\,n_p+1}^p - \tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}{\displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \tag{41}$$

or, equivalently,

$$T_2 = \zeta_{im}^{(n_c,\,n_p)} \left[\frac{\tau_{i,\,n_p+1}^p - \tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}{\tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right]. \tag{42}$$

This is readily seen to be

$$T_2 = \zeta_{im}^{(n_c,\,n_p)} \left[1 - \frac{\tau_{i,\,n_p+1}^p}{\tau_{i,\,n_p+1}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right], \tag{43}$$

which leads to the final result for a partially categorized observation

$$\zeta_{im}^{(n_c, n_p + 1)} = \zeta_{im}^{(n_c, n_p)} + \left[\frac{\tau_{i, n_p + 1}^p}{\tau_{i, n_p + 1}^p + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\xi_{jm} - \zeta_{im}^{(n_c, n_p)} \right]. \qquad (44)$$

Recall that no update takes place for an uncategorized observation.

Recursive Update Equation for π

Consider first $\pi_i^{(n_c + 1, n_p, n_u)}$, which implies that the latest observation is class categorized. In terms of the E-M expression,

$$\pi_i^{(n_c + 1, n_p, n_u)} = \left[\frac{\sum_{j=1}^{n_c + 1} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{(n_c + 1) + n_p + n_u} \right] \qquad (45)$$

or, equivalently

$$\pi_i^{(n_c + 1, n_p, n_u)} = \left[\frac{\tau_{i(n_c + 1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{(n_c + 1) + n_p + n_u} \right]. \qquad (46)$$

The right-hand side of this equation can be broken into two terms, corresponding to the last observation and all previous observations, respectively.

$$\pi_i^{(n_c + 1, n_p, n_u)} = \frac{\tau_{i(n_c + 1)}^c}{(n_c + 1) + n_p + n_u} + \left[\frac{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{(n_c + 1) + n_p + n_u} \right] \qquad (47)$$

This can be rewritten as

$$\pi_i^{(n_c + 1, n_p, n_u)} = \frac{\tau_{i(n_c + 1)}^c}{(n_c + 1) + n_p + n_u} + \left[\frac{n_c + n_p + n_u}{(n_c + 1) + n_p + n_u} \right] \pi_i^{(n_c, n_p, n_u)}, \tag{48}$$

which becomes

$$\pi_i^{(n_c + 1, n_p, n_u)} = \pi_i^{(n_c, n_p, n_u)} + \frac{1}{(n_c + n_p + n_u) + 1} \left[\tau_{i(n_c + 1)}^c - \pi_i^{(n_c, n_p, n_u)} \right]. \tag{49}$$

Consider next $\pi_{im}^{(n_c, n_p + 1, n_u)}$, which implies that the latest observation is partially class categorized. In terms of the E-M expression,

$$\pi_i^{(n_c, n_p + 1, n_u)} = \left[\frac{\displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p + 1} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{n_c + (n_p + 1) + n_u} \right] \tag{50}$$

or, equivalently

$$\pi_i^{(n_c, n_p + 1, n_u)} = \left[\frac{\tau_{i(n_p + 1)}^p + \displaystyle\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{n_c + (n_p + 1) + n_u} \right]. \tag{51}$$

This readily becomes

$$\pi_i^{(n_c, n_p + 1, n_u)} = \frac{\tau_{i(n_p + 1)}^p}{n_c + (n_p + 1) + n_u} + \left[\frac{n_c + n_p + n_u}{n_c + (n_p + 1) + n_u} \right] \pi_i^{(n_c, n_p, n_u)}, \tag{52}$$

which can be rewritten as

$$\pi_i^{(n_c, n_p + 1, n_u)} = \pi_i^{(n_c, n_p, n_u)} + \frac{1}{(n_c + n_p + n_u) + 1} \left[\tau_{i(n_p + 1)}^p - \pi_i^{(n_c, n_p, n_u)} \right]. \tag{53}$$

For uncategorized observations, it is easy to show that the correct expression is

$$\pi_i^{(n_c, n_p, n_u + 1)} = \pi_i^{(n_c, n_p, n_u)} + \frac{1}{(n_c + n_p + n_u) + 1} \left[\tau_{i(n_u + 1)}^u - \pi_i^{(n_c, n_p, n_u)} \right].$$ (54)

Recursive Update Equation for μ

Consider first $\mu_i^{(n_c + 1, n_p, n_u)}$, which implies that the latest observation is class categorized. In terms of the E-M expression,

$$\mu_i^{(n_c + 1, n_p, n_u)} = \left[\frac{\sum\limits_{j=1}^{n_c + 1} \left[\tau_{ij}^c x_j \right] + \sum\limits_{j=1}^{n_p} \left[\tau_{ij}^p x_j \right] + \sum\limits_{j=1}^{n_u} \left[\tau_{ij}^u x_j \right]}{\sum\limits_{j=1}^{n_c + 1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \right]$$ (55)

or, equivalently

$$\mu_i^{(n_c + 1, n_p, n_u)} = \left[\frac{\tau_{i(n_c + 1)}^c x_{n_c + 1} + \sum\limits_{j=1}^{n_c} \left[\tau_{ij}^c x_j \right] + \sum\limits_{j=1}^{n_p} \left[\tau_{ij}^p x_j \right] + \sum\limits_{j=1}^{n_u} \left[\tau_{ij}^u x_j \right]}{\sum\limits_{j=1}^{n_c + 1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \right].$$ (56)

The right-hand side of this equation can be broken into two terms, corresponding to the last observation and all previous observations, respectively.

$$\mu_i^{(n_c + 1, n_p, n_u)} = \frac{\tau_{i(n_c + 1)}^c x_{n_c + 1}}{\sum\limits_{j=1}^{n_c + 1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} + \left[\frac{\sum\limits_{j=1}^{n_c} \left[\tau_{ij}^c x_j \right] + \sum\limits_{j=1}^{n_p} \left[\tau_{ij}^p x_j \right] + \sum\limits_{j=1}^{n_u} \left[\tau_{ij}^u x_j \right]}{\sum\limits_{j=1}^{n_c + 1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \right]$$ (57)

This can be rewritten as

$$\mu_i^{(n_c+1,\,n_p,\,n_u)} = \frac{\tau_{i\,(n_c+1)}^c\, x_{n_c+1}}{\displaystyle\sum_{j=1}^{n_c+1}\tau_{ij}^c + \sum_{j=1}^{n_p}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u} + \left[\frac{\displaystyle\sum_{j=1}^{n_c}\tau_{ij}^c + \sum_{j=1}^{n_p}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u}{\displaystyle\sum_{j=1}^{n_c+1}\tau_{ij}^c + \sum_{j=1}^{n_p}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u}\right]\mu_i^{(n_c,\,n_p,\,n_u)}\,, \qquad (58)$$

which becomes

$$\mu_i^{(n_c+1,\,n_p,\,n_u)} = \mu_i^{(n_c,\,n_p,\,n_u)} + \frac{\tau_{i\,(n_c+1)}^c}{\displaystyle\sum_{j=1}^{n_c+1}\tau_{ij}^c + \sum_{j=1}^{n_p}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u}\left[x_{n_c+1} - \mu_i^{(n_c,\,n_p,\,n_u)}\right]. \qquad (59)$$

Consider next $\mu_{im}^{(n_c,\,n_p+1,\,n_u)}$, which implies that the latest observation is partially class categorized. In terms of the E-M expression,

$$\mu_i^{(n_c,\,n_p+1,\,n_u)} = \left[\frac{\displaystyle\sum_{j=1}^{n_c}\left[\tau_{ij}^c x_j\right] + \sum_{j=1}^{n_p+1}\left[\tau_{ij}^p x_j\right] + \sum_{j=1}^{n_u}\left[\tau_{ij}^u x_j\right]}{\displaystyle\sum_{j=1}^{n_c}\tau_{ij}^c + \sum_{j=1}^{n_p+1}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u}\right] \qquad (60)$$

or, equivalently

$$\mu_i^{(n_c,\,n_p+1,\,n_u)} = \left[\frac{\tau_{i\,(n_p+1)}^p + \displaystyle\sum_{j=1}^{n_c}\left[\tau_{ij}^c x_j\right] + \sum_{j=1}^{n_p}\left[\tau_{ij}^p x_j\right] + \sum_{j=1}^{n_u}\left[\tau_{ij}^u x_j\right]}{\displaystyle\sum_{j=1}^{n_c}\tau_{ij}^c + \sum_{j=1}^{n_p+1}\tau_{ij}^p + \sum_{j=1}^{n_u}\tau_{ij}^u}\right], \qquad (61)$$

this readily becomes

$$\mu_i^{(n_c, n_p + 1, n_u)} = \frac{\tau_{i(n_p + 1)}^p}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p + 1} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} + \left[\frac{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p + 1} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} \right] \mu_i^{(n_c, n_p, n_u)} , \qquad (62)$$

which becomes

$$\mu_i^{(n_c, n_p + 1, n_u)} = \mu_i^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_p + 1)}^p}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p + 1} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} \left[x_{i(n_p + 1)} - \mu_i^{(n_c, n_p, n_u)} \right] . \qquad (63)$$

For uncategorized observations it is easy to show that the correct expression is

$$\mu_i^{(n_c, n_p, n_u + 1)} = \mu_i^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_u + 1)}^u}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u + 1} \tau_{ij}^u} \left[x_{i(n_u + 1)} - \mu_i^{(n_c, n_p, n_u)} \right] . \qquad (64)$$

Recursive Update Equation for Σ

Consider first $\Sigma_i^{(n_c + 1, n_p, n_u)}$, which implies that the latest observation is class categorized. The E-M expression is

$$\Sigma_{ii}^{(n_c + 1, n_p, n_u)} = \left[\frac{\sum_{j=1}^{n_c + 1} \left[\tau_{ij}^c (x_j - \mu_i)^2 \right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^p (x_j - \mu_i)^2 \right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^u (x_j - \mu_i)^2 \right]}{\sum_{j=1}^{n_c + 1} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} \right] \qquad (65)$$

or, equivalently

$$\Sigma_{ii}^{(n_c+1, n_p, n_u)} = \left[\frac{\tau_{i(n_c+1)}^{c} (x_{n_c+1} - \mu_i)^2}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} \right]$$

$$+ \left[\frac{\sum_{j=1}^{n_c} \left[\tau_{ij}^{c} (x_j - \mu_i)^2 \right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^{p} (x_j - \mu_i)^2 \right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^{u} (x_j - \mu_i)^2 \right]}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} \right]. \quad (66)$$

This can be rewritten as

$$\Sigma_{ii}^{(n_c+1, n_p, n_u)} = \frac{\tau_{i(n_c+1)}^{c} (x_{n_c+1} - \mu_i)^2}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} + \left[\frac{\sum_{j=1}^{n_c} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} \right] \Sigma_{ii}^{(n_c, n_p, n_u)}, \quad (67)$$

which becomes

$$\Sigma_{ii}^{(n_c+1, n_p, n_u)} = \Sigma_{ii}^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_c+1)}^{c}}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} \left[(x_{n_c+1} - \mu_i)^2 - \Sigma_{ii}^{(n_c, n_p, n_u)} \right]. \quad (68)$$

Similarly, for partially categorized and uncategorized data, respectively,

$$\Sigma_{ii}^{(n_c, n_p+1, n_u)} = \Sigma_{ii}^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_p+1)}^{p}}{\sum_{j=1}^{n_c+1} \tau_{ij}^{c} + \sum_{j=1}^{n_p+1} \tau_{ij}^{p} + \sum_{j=1}^{n_u} \tau_{ij}^{u}} \left[(x_{n_p+1} - \mu_i)^2 - \Sigma_{ii}^{(n_c, n_p, n_u)} \right] \quad (69)$$

and

$$\Sigma_{ii}^{(n_c, n_p, n_u + 1)} = \Sigma_{ii}^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_u + 1)}^u}{\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u + 1} \tau_{ij}^u} \left[(x_{n_u + 1} - \mu_i)^2 - \Sigma_{ii}^{(n_c, n_p, n_u)} \right]. \quad (70)$$

RESULTS

SUMMARY OF UNIVARIATE ITERATIVE E-M UPDATE EQUATIONS

For completeness, the iterative E-M equations are reproduced first.

$$\zeta_{im} = \left[\frac{1}{\sum_{j=1}^{n_c} \sum_{m=1}^{M} \tau_{ijm} + \sum_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\sum_{j=1}^{n_c} \tau_{ijm} + \sum_{j=1}^{n_p} \xi_{jm} \tau_{ij}^p \right], \quad (71)$$

$$\pi_i = \frac{1}{(n_c + n_p + n_u)} \left[\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u \right], \quad (72)$$

$$\mu_i = \frac{\sum_{j=1}^{n_c} \left[\tau_{ij}^c x_j \right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^p x_j \right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^u x_j \right]}{\left[\sum_{j=1}^{n_c} \left[\tau_{ij}^c \right] + \sum_{j=1}^{n_p} \left[\tau_{ij}^p \right] + \sum_{j=1}^{n_u} \left[\tau_{ij}^u \right] \right]}, \quad (73)$$

$$\Sigma_{ii} = \frac{\sum_{j=1}^{n_c} \tau_{ij}^c (x_j - \mu_i)^2 + \sum_{j=1}^{n_p} \tau_{ij}^p (x_j - \mu_i)^2 + \sum_{j=1}^{n_u} \tau_{ij}^u (x_j - \mu_i)^2}{\left[\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u \right]}, \quad (74)$$

where

$$\tau_{ijm} = \frac{z_{jm}\zeta_{im}\pi_i f(x_j;\theta_i)}{\displaystyle\sum_{i=1}^{g}\sum_{m=1}^{M} z_{jm}\zeta_{im}\pi_i f(x_j;\theta_i)} \qquad (75)$$

and

$$\tau_{ij}^{c} = \sum_{m=1}^{M} \tau_{ijm} = \frac{\zeta_{im_j}\pi_i f(x_j;\theta_i)}{\displaystyle\sum_{i=1}^{g}\zeta_{im_j}\pi_i f(x_j;\theta_i)} = \frac{\displaystyle\sum_{m=1}^{M} z_{jm}\zeta_{im}\pi_i f(x_j;\theta_i)}{\displaystyle\sum_{i=1}^{g}\sum_{m=1}^{M} z_{jm}\zeta_{im}\pi_i f(x_j;\theta_i)} . \qquad (76)$$

Similarly

$$\tau_{ij}^{p} = \tau_{ij}^{u} = \frac{\pi_i f(x_j;\theta_i)}{\displaystyle\sum_{i=1}^{g}\pi_i f(x_j;\theta_i)} , \qquad (77)$$

where it is to be remembered that τ_{ij}^{p} is only computed when x_j is a partially categorized observation and similarly for τ_{ij}^{u}.

The E-M algorithm then consists of iterating the expectation step consisting of evaluating Equations (75), (76), and (77) for the appropriate observations and the maximization step, which consists of evaluating new parameter values using Equations (71) through (74).

The multivariate versions can be obtained by making x_j and μ_i vector quantities and Σ_i a matrix. The equations for μ_i and Σ_i become

$$\mu_i^k = \frac{\displaystyle\sum_{i=1}^{n_c}\left[\tau_{ij}^c x_j^k\right] + \sum_{j=1}^{n_p}\left[\tau_{ij}^p x_j^k\right] + \sum_{j=1}^{n_u}\left[\tau_{ij}^u x_j^k\right]}{\left[\displaystyle\sum_{j=1}^{n_c}\left[\tau_{ij}^c\right] + \sum_{j=1}^{n_p}\left[\tau_{ij}^p\right] + \sum_{j=1}^{n_u}\left[\tau_{ij}^u\right]\right]} \qquad (78)$$

and

19

$$\Sigma_i^{kl} = \frac{\sum_{j=1}^{n_c} \tau_{ij}^c \left(x_j^k - \mu_i^k \right) \left(x_j^l - \mu_i^l \right) + \sum_{j=1}^{n_p} \tau_{ij}^p \left(x_j^k - \mu_i^k \right) \left(x_j^l - \mu_i^l \right) + \sum_{j=1}^{n_u} \tau_{ij}^u \left(x_j^k - \mu_i^k \right) \left(x_j^l - \mu_i^l \right)}{\left[\sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u \right]}, \quad (79)$$

where the component indices are denoted by superscripts.

SUMMARY OF UNIVARIATE RECURSIVE UPDATE EQUATIONS

Categorized Observation

$$\zeta_{im}^{(n_c+1, n_p)} = \zeta_{im}^{(n_c, n_p)} + \frac{\tau_{i(n_c+1)}^c \cdot \left[z_{(n_c+1)m} - \zeta_{im}^{(n_c, n_p)} \right]}{\tau_{i(n_c+1)}^c + \sum_{j=1}^{n_c} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p}. \quad (80)$$

$$\pi_i^{(n_c+1, n_p, n_u)} = \pi_i^{(n_c, n_p, n_u)} + \frac{1}{(n_c + n_p + n_u) + 1} \left[\tau_{i(n_c+1)}^c - \pi_i^{(n_c, n_p, n_u)} \right]. \quad (81)$$

$$\mu_i^{(n_c+1, n_p, n_u)} = \mu_i^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_c+1)}^c}{\sum_{j=1}^{n_c+1} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} \left[x_{n_c+1} - \mu_i^{(n_c, n_p, n_u)} \right]. \quad (82)$$

$$\Sigma_{ii}^{(n_c+1, n_p, n_u)} = \Sigma_{ii}^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_c+1)}^c}{\sum_{j=1}^{n_c+1} \tau_{ij}^c + \sum_{j=1}^{n_p} \tau_{ij}^p + \sum_{j=1}^{n_u} \tau_{ij}^u} \left[\left(x_{n_c+1} - \mu_i \right)^2 - \Sigma_{ii}^{(n_c, n_p, n_u)} \right]. \quad (83)$$

For the multivariate case, Equations (82) and (83) become (with vector indices k and l)

$$\mu_i^{k\,(n_c+1,\,n_p,\,n_u)} = \mu_i^{k\,(n_c,\,n_p,\,n_u)} + \frac{\tau_{i\,(n_c+1)}^c}{\sum\limits_{j=1}^{n_c+1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[x_{n_c+1}^k - \mu_i^{k\,(n_c,\,n_p,\,n_u)} \right], \qquad (84)$$

$$\Sigma_i^{kl\,(n_c+1,\,n_p,\,n_u)} = \Sigma_i^{kl\,(n_c,\,n_p,\,n_u)}$$

$$+ \frac{\tau_{i\,(n_c+1)}^c}{\sum\limits_{j=1}^{n_c+1} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[\left(x_{n_c+1}^k - \mu_i^k \right) \left(x_{n_c+1}^l - \mu_i^l \right) - \Sigma_i^{kl\,(n_c,\,n_p,\,n_u)} \right]. \qquad (85)$$

Partially Categorized Observation

$$\zeta_{im}^{(n_c,\,n_p+1)} = \zeta_{im}^{(n_c,\,n_p)} + \left[\frac{\tau_{i,\,n_p+1}^p}{\tau_{i,\,n_p+1}^p + \sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p} \right] \left[\xi_{jm} - \zeta_{im}^{(n_c,\,n_p)} \right]. \qquad (86)$$

$$\pi_i^{(n_c,\,n_p+1,\,n_u)} = \pi_i^{(n_c,\,n_p,\,n_u)} + \frac{1}{(n_c+n_p+n_u)+1} \left[\tau_{i\,(n_p+1)}^p - \pi_i^{(n_c,\,n_p,\,n_u)} \right]. \qquad (87)$$

$$\mu_i^{(n_c,\,n_p+1,\,n_u)} = \mu_i^{(n_c,\,n_p,\,n_u)} + \frac{\tau_{i\,(n_p+1)}^p}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p+1} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[x_{n_p+1} - \mu_i^{(n_c,\,n_p,\,n_u)} \right]. \qquad (88)$$

$$\Sigma_{ii}^{(n_c,\,n_p+1,\,n_u)} = \Sigma_{ii}^{(n_c,\,n_p,\,n_u)} + \frac{\tau_{i\,(n_p+1)}^p}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p+1} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[\left(x_{n_p+1} - \mu_i \right)^2 - \Sigma_{ii}^{(n_c,\,n_p,\,n_u)} \right]. \qquad (89)$$

For the multivariate case, Equations (88) and (89) become (with vector indices k and l)

$$\mu_i^{k(n_c, n_p + 1, n_u)} = \mu_i^{k(n_c, n_p, n_u)} + \frac{\tau_{i(n_p+1)}^p}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p+1} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[x_{n_p+1}^k - \mu_i^{k(n_c, n_p, n_u)} \right], \tag{90}$$

$$\Sigma_i^{kl(n_c, n_p + 1, n_u)} = \Sigma_i^{kl(n_c, n_p, n_u)}$$

$$+ \frac{\tau_{i(n_p+1)}^p}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p+1} \tau_{ij}^p + \sum\limits_{j=1}^{n_u} \tau_{ij}^u} \left[\left(x_{n_p+1}^k - \mu_i^k \right) \left(x_{n_p+1}^l - \mu_i^l \right) - \Sigma_i^{kl(n_c, n_p, n_u)} \right]. \tag{91}$$

Uncategorized Observation

$$\pi_i^{(n_c, n_p, n_u + 1)} = \pi_i^{(n_c, n_p, n_u)} + \frac{1}{(n_c + n_p + n_u) + 1} \left[\tau_{i(n_u+1)}^u - \pi_i^{(n_c, n_p, n_u)} \right]. \tag{92}$$

$$\mu_i^{(n_c, n_p, n_u + 1)} = \mu_i^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_u+1)}^u}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u+1} \tau_{ij}^u} \left[x_{n_u+1} - \mu_i^{(n_c, n_p, n_u)} \right]. \tag{93}$$

$$\Sigma_{ii}^{(n_c, n_p, n_u + 1)} = \Sigma_{ii}^{(n_c, n_p, n_u)} + \frac{\tau_{i(n_u+1)}^u}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u+1} \tau_{ij}^u} \left[(x_{n_u+1} - \mu_i)^2 - \Sigma_{ii}^{(n_c, n_p, n_u)} \right]. \tag{94}$$

For the multivariate case, Equations (88) and (89) become (with vector indices k and l)

$$\mu_i^{k(n_c, n_p, n_u + 1)} = \mu_i^{k(n_c, n_p, n_u)} + \frac{\tau_{i(n_u+1)}^u}{\sum\limits_{j=1}^{n_c} \tau_{ij}^c + \sum\limits_{j=1}^{n_p} \tau_{ij}^p + \sum\limits_{j=1}^{n_u+1} \tau_{ij}^u} \left[x_{n_u+1}^k - \mu_i^{k(n_c, n_p, n_u)} \right], \tag{95}$$

22

$$\Sigma_i^{kl\,(n_c,\,n_p,\,n_u+1)} = \Sigma_i^{kl\,(n_c,\,n_p,\,n_u)}$$

$$+ \frac{\tau_{i\,(n_u+1)}^u}{\displaystyle\sum_{j=1}^{n_c}\tau_{ij}^c + \sum_{j=1}^{n_p}\tau_{ij}^p + \sum_{j=1}^{n_u+1}\tau_{ij}^u} \left[\left(x_{n_u+1}^k - \mu_i^k\right)\left(x_{n_u+1}^l - \mu_i^l\right) - \Sigma_i^{kl\,(n_c,\,n_p,\,n_u)}\right]. \tag{96}$$

CONCLUSIONS

The derivation of the recursive E-M equations for joint representation mixture models with normal components has been presented. While the detailed derivation was for the univariate case, a slightly more complicated derivation results in the full multivariate equations. The results for this case have been presented without derivation.

The joint representation approach represents a significant philosophical departure from current mixture model usage. The standard mixture model usage is either to build a separate mixture model for each class when the observations are class labeled, or to assume that each class is normally distributed so that a mixture model for all the data can be interpreted as a mixture of normal classes. This approach, in effect, totally relaxes the requirement for each class to be normally distributed. Philosophically, a semiparametric viewpoint has been taken in that it is assumed that each class can be modelled by a (potentially complex) mixture model and that no significance is to be ascribed to an individual term in the mixture. As an example, contrast a mixture model approximation to a lognormal density to a mixture of two normals. In the latter case, it may well make sense to care about which of the two terms gave rise to a particular observation. However, in the lognormal case, where by assumption the density is nonparametric with respect to representation by normal mixtures, this sort of distinction has little or no meaning.

This approach is thus appropriate for combined supervised/unsupervised (various levels of class categorization) learning when the individual class densities may be more complex than simple normals. It provides a unified framework for handling this problem. Once the joint representation density has been estimated, densities corresponding to the individual classes can be easily recovered.

The recursive versions of these equations allow this approach to be used for large data sets as well as in an adaptive mixture model framework.

23

REFERENCES

1.   Titterington, D. M.; Smith, A. F. M.; and Makov, U. E., *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.

2.   McLachlan, G. J. and Basford, K. E., *Mixture Models*, Marcel Dekker, New York, 1988.

3.   Priebe, C. E., "Adaptive Mixtures", *J. Am. Statist. Assoc.*, Vol. 89, 1994, pp. 796-806.

4.   Rogers, G. W.; Marchette, D. J.; and Priebe, C. E., "A Procedure for Model Complexity Selection in Semiparametric Mixture Model Density Estimation," *Math. Modeling and Scientific Comp.*, to appear (1995).

5.   Solka, J. L., *Matching Model Information Content to Data Information*, Ph.D. Thesis, George Mason Univ., 1995.

6.   Priebe, C. E.; Julin, E. G.; Rogers, G. W.; Healy, D. M.; Lu , J.; Solka, J. L.; and Marchette, D. J., "Incorporating Segmentation Boundaries into the Calculation of Fractal Dimension Features," *Proc. 26th Symp. on the Interface*, Research Triangle Park, NC, 15-18 Jun 1994, pp. 52-56.

7.   Priebe, C. E.; Lorey, R. A.; Marchette, D. J.; Solka, J. L.; and Rogers, G. W., "Nonparametric Spatio-temporal Change Point Analysis for Early Detection in Mammography," *Proc. of the Second Int'l Workshop on Digital Mammography (SIWDM)*, York, UK, 10-12 Jul 1994, pp. 111-120.

8.   Priebe, C. E.; Rogers, G. W.; Marchette, D. J.; and Solka, J. L., "Change Point Analysis With Adaptive Mixture Models," *Proc. of the Int'l Geoscience and Remote Sensing Symp. (IGARSS '94)*, Pasadena, CA, 8-12 Aug 1994, pp. 1795-1798.

9.   Solka, J. L.; Priebe, C. E.; Rogers, G. W.; Poston, W. L.; and Lorey, R. A., "Maximum Likelihood Density Estimation with Term Creation and Annihilation," *Proc. 26th Symp. on the Interface*, Research Triangle Park, NC, 15-18 Jun 1994, pp. 222-225.

10.   Priebe, C. E.; Solka, J. L.; Lorey, R. A.; Rogers, G. W.; Poston, W. L.; Kallergi, M.; Qian,W.; Clarke, L. P.; and Clark, R. A., "The Application of Fractal Analysis to Mammographic Tissue Classification," *Cancer Letters*, Vol. 77, 1994, pp. 183-189.

11.   Solka, J. L.; Priebe, C. E.; and Rogers, G. W., "An Initial Assessment of Discriminant Surface Complexity for Power Law Features," *Simulation: Special Issue: Neural Networks: Model Development or Applications*, Vol. 58, 1992, pp. 311-318.

REFERENCES (Continued)

12.	Rogers, G. W.; and Lorey, R. A., *Derivation of the Joint Representation Mixture Model Equations*, NSWCDD/TR-95/92, May 1995, Dahlgren, Virginia.

DISTRIBUTION

<u>COPIES</u>

<u>COPIES</u>

DOD ACTIVITIES (CONUS)

B10	(SOLKA)	1
C		1
D		1
E231		3
E282	(SWANSBURG)	1
G33	(POSTON)	1

ATTN   CODE 342PS (HAWKINS)   1
OFFICE OF NAVAL RESEARCH
800 N QUINCY ST
ARLINGTON VA 22217

ATTN   CODE E29L
              (TECHNICAL LIBRARY)   1
COMMANDING OFFICER
CSSDD NSWC
6703 W HIGHWAY 98
PANAMA CITY FL 32407-7001

DEFENSE TECHNICAL INFORMATION CTR
8725 JOHN J KINGMAN
SUITE 0944
FT BELVOIR VA 22060-6218   2

NON-DOD ACTIVITIES (CONUS)

THE CNA CORPORATION
PO BOX 16268
ALEXANDRIA VA 22302-0268   1

ATTN   GIFT AND EXCHANGE DIVISION   4
LIBRARY OF CONGRESS
WASHINGTON DC 20540

ATTN   PRIEBE   1
DEPT OF MATHEMATICAL SCIENCES
JOHNS HOPKINS UNIVERSITY
BALTIMORE MD 21218

INTERNAL

B		1
B05	(MOORE)	1
B10	(ROGERS)	10
B10	(LOREY)	1
B10	(MARCHETTE)	1